

I.1

The Rice Genome Sequence as an Indispensable Tool for Crop Improvement

Takuji Sasaki¹(✉), Jianzhong Wu¹, Hiroshi Mizuno¹, Baltazar A. Antonio¹, and Takashi Matsumoto¹

1 Introduction

Rice is one of the three major staple food crops in the world. In 2004, world rice production representing total rice yield was about 450 million tons. This indicates that the introduction of high-yielding rice varieties coupled with improvements in agricultural practices over the last three decades have made a major impact in the form of increased rice production. However, although the global production and consumption of rice is currently at an equilibrium (<http://worldfood.apionet.or.jp/index-e.html>), the majority of the population in rice-producing areas, particularly in many Asian and African countries, is still suffering from hunger, malnutrition and extreme poverty. Although a stable supply of rice is closely associated with the agricultural policies of each country and the natural environment that supports agricultural activities, the scientific community can play a major role in maintaining a sustainable agriculture system for the benefit of mankind. Therefore, continuous efforts towards innovative research should be encouraged in order to address the many aspects related to increasing rice productivity in the midst of all the obstacles associated with rice cultivation, such as land scarcity, depleted water resources and climatic changes. Fortunately, recent advances in the field of genomics may offer new opportunities to tackle problems associated with crop improvement and agricultural productivity.

Characterization and sequencing of the genome has been carried out in many organisms including plants as well as many scientifically and practically important eukaryotes. This approach has led to a thorough analysis of the relationship between gene activity and the genomic structure of each target species that characterizes it. It has also facilitated the identification of genes corresponding to phenotypes, which differentiates one species from another within the same genus or other closely related genera. Genomics has already revolutionized many aspects of basic research which has resulted in many surprising biological discoveries not only in

¹National Institute of Agrobiological Sciences, 1–2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan
e-mail: tsasaki@nias.affrc.go.jp

agriculture, but also in medicine and other biologically based industries. In the case of major cereal crops such as rice and maize, large-scale genome analyses have been carried out since the early 1990s with the aim of identifying genes practically important to agriculture. These include genes related to yield potential, biotic and abiotic stress tolerance, heading date, and other agronomically important traits which are either orthologous among cereal crops or unique to each crop.

Rice is the first cereal crop to be completely decoded (International Rice Genome Sequencing Project 2005). The high-quality map-based sequence of the entire rice genome is now available in the public domain. With the completion of sequencing, the next challenge to the scientific community would be to determine the function of about 37,000 predicted genes in rice and to utilize this information in identifying agronomically important genes not only in cultivated rice, but also in wild relatives of rice and in other cereal crops.

2 Major Features of the Rice Genome

The *japonica* rice variety “Nipponbare” was chosen as a common template for complete and accurate decoding of the *Oryza sativa* genome. This is because Nipponbare was one of the parent cultivars of the F2 population used for linkage analysis to construct a high-density molecular genetic map of rice (Harushima et al. 1998). The same cultivar was also used as a resource for the construction of cDNA libraries in order to generate a catalog of rice ESTs. About 6,000 of these ESTs were used as genetic and/or PCR-based markers to reconstruct the rice genome using large-sized rice genomic DNA fragments ligated in yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), and P1-derived artificial chromosome (PAC) vectors. Additionally, Nipponbare could be easily regenerated from callus, an important advantage for genetic transformation.

The International Rice Genome Sequencing Project (IRGSP), a consortium of publicly funded laboratories from ten countries, was formed in 1998 to pursue the sequencing of the genome. In 2004, the IRGSP succeeded in completely decoding the rice genome sequence with 99.99% accuracy (International Rice Genome Sequencing Project 2005). The major features of the rice genome based on the high-quality genome sequence are as follows:

1. The genome size of Nipponbare is 389 Mb including the unsequenced regions or gaps in the physical map, which were measured by fiber-FISH and conventional FISH.
2. The genome consists of 37,544 non-transposable element-related genes, which were predicted *ab initio* by FGENESH after masking the repeat sequences.
3. Twenty nine percent of the genes (10,837) are amplified at least once in tandem within an adjacent 5 Mb region of the genome.
4. Chloroplast and mitochondrial insertions contribute 0.20–0.24% and 0.18–0.19% of the nuclear genome, respectively.

5. Class I and class II transposons occupy 19.4% and 13% of the rice genome, respectively.
6. A total of 18,828 class 1 simple sequence repeats (SSR) were identified as di, tri, and tetra-nucleotide SSRs.

A detailed annotation of the completed rice genome sequence was carried out using rice full-length cDNAs as basic tools. This effort resulted in the accurate prediction of 29,550 expressed loci including the unmapped-mRNA clusters (The Rice Annotation Project, 2007; <http://rapdb.dna.affrc.go.jp/>). The details of this annotation and curation system are described by Ito et al. in Chapter I.2 of this book.

2.1 Characterization of Rice Centromeres

Rice centromeres have been characterized by the presence of a highly repetitive 155–165 bp satellite DNA, CentO, along with centromere-specific retrotransposons. Using CentO as a probe, the size of the centromeres of chromosome 4 and 8 were estimated cytogenetically as 50–100 kb (Cheng et al. 2002). Therefore, physical mapping of these genomic regions by assignment of fingerprint contigs using DNA markers as anchors and chromosome walking strategies has been a major problem. The structures of the centromeres of chromosomes 4 and 8 were found to be strikingly different from each other. In the case of the 124 kb physical map of chromosome 4 centromere, 18 tracts of 379 tandemly arrayed CentO units were identified (Zhang et al. 2004). On the other hand, in the case of chromosome 8 centromere, the core region of 30 kb consisted of only three tracts of 452 CentO units. The genomic regions between each CentO consisted of retroelements such as RIRE7, solo LTR, or Ty3-gypsy (Wu et al. 2004; Ma and Bennetzen 2005). Although there is no reasonable explanation based on biological function or the divergence and/or evolution of the centromeres, the dynamic amplification and/or recombination of centromeric repeats and accumulation of retrotransposons in the centromere could account for the difference between chromosome 4 and 8 centromeres (Ma and Jackson 2007; Ma et al. 2007). A comparative analysis of the centromere structure among the twelve chromosomes in several rice ecotypes should be carried out to address this phenomenon (Mizuno et al. 2007).

A detailed genomic characterization has also been performed on the centromere of chromosome 3. The size of the CentO-rich core region, however, was too large (about 480 kb), which makes it unrealistic to fully sequence. Therefore, cytogenetic analysis by fiber-FISH using CentO as a probe was performed to reveal several long CentO clusters interrupted by non-CentO blocks within the core. The resolution of fiber-FISH allowed for further interpretation of the arrangement of CentO of chromosome 3 (The Rice Chromosome 3 Sequencing Consortium, 2005). The centromere structure of chromosome 5 has also been recently elucidated (unpublished). The physical map of chromosome 5 includes two PAC clones covering the centromere with 4 kb and 65 kb CentO clusters in the sequence and shows similar structural characteristics to that of chromosome 8.

3 Comparative Analysis of Genome Structure Within *Oryza sativa*

The genus *Oryza* includes more than 120,000 varieties cultivated throughout the world, as well as many wild rice species. The cultivar Nipponbare belongs to *Oryza sativa* ssp. *japonica*, which is mainly adapted to temperate regions and cultivated in limited areas such as Japan, Korea, Taiwan, and some parts of Europe. The subspecies *indica*, on the other hand, is widely cultivated in many countries, including China, India, Thailand, Indonesia, Senegal, and the USA. Therefore, *indica* rice has a wider range of genetic diversity compared with *japonica* rice. Furthermore, *indica* rice varieties account for about 90% of total rice production worldwide. This suggests that the genome sequence information of an *indica* rice variety is just as desirable and important in agriculture as the genome sequence of a *japonica* rice variety. Phylogenetic analysis has shown the existence of subgroups among *indica* rice varieties (Garris et al. 2005). So far, the genome sequence of 93-11, an *indica* rice variety cultivated in China, has been obtained by whole-genome shotgun sequencing (Yu et al. 2002). The assembled draft sequence totals 466 Mb, which is much larger than the estimated genome size for Nipponbare. Since it is only a draft sequence, sequence fragments of transposable elements may have been redundantly incorporated into nucleotide sequences of euchromatic regions. To confirm the actual genome size of 93-11 would require a similar map-based sequencing strategy as performed in Nipponbare.

The difference in genome size between the subspecies of cultivated rice was clarified by *in silico* physical mapping of *indica* rice variety Kasalath, using the Nipponbare physical map as a standard (Katagiri et al. 2004). A BAC library of Kasalath was constructed and the end-sequence of each BAC clone was analyzed. After removing the BACs carrying repetitive or unrealistically distant pairs of end sequences, each pair of end sequences was blasted to the Nipponbare genome sequence. As a result, a total of 12,170 paired BACs were mapped to the 12 chromosomes, representing 450 contigs and a total length of 308.5 Mb. To further confirm the alignment of Kasalath BACs along the Nipponbare chromosome 1, PCR primers were designed based on the Nipponbare EST sequences and used for mapping the Kasalath BACs. Of the 306 Kasalath BACs corresponding to the minimum tiling path of chromosome 1, a total of 290 BACs were confirmed as containing the marker sequences derived from expressed genes. The mapping results indicate that the difference in the genome size of the two rice subspecies is rather small. However, the frequency of Kasalath BACs mapped in chromosomes 4, 9, 11, and 12 was relatively low. In the case of chromosomes 4 and 9, the low recovery could be attributed to a large amount of heterochromatin in these chromosomes. Detailed sequence comparison of a 2.3 Mb region of chromosome 4 between Nipponbare and a Chinese *indica* variety, Guangluai 4, revealed the colinearity of gene order and more insertions in *japonica* than in *indica* (Feng et al. 2002). On the other hand, in the case of chromosomes 11 and

12, the low frequency of mapped Kasalath BAC clones recovery may be due to the difference in the genome structure of Nipponbare and Kasalath. A 1.35 Mb region in chromosome 11 and a 0.69 Mb region in chromosome 12 of Nipponbare did not generate corresponding Kasalath BAC contigs. In chromosome 11, this region was found to be rich in disease resistance genes and varied from one cultivar to another (Leister et al. 1998).

To facilitate the identification of allelic divergence among *O. sativa*, the genetic diversity among 30,000 rice germplasm accessions at the Gene Bank of the National Institute of Agrobiological Sciences (NIAS) in Japan was evaluated. After the first screening based on the available data such as regional origin and morphological appearance of each accession, 332 accessions were selected and evaluated for polymorphism using 179 loci of RFLPs (Kojima et al. 2005). The RFLP data were subjected to cluster analysis and 67 groups were recognized. A single accession from each of the 67 groups was selected. These 67 accessions retained 91% of the alleles in the original 332 accessions. This core collection is now available at the NIAS Gene Bank (http://www.gene.affrc.go.jp/index_j.php) and can be used as an important resource for detailed genetic studies and for further improvement of *O. sativa* cultivars.

4 Comparative Analysis of Genome Structure in the Genus *Oryza*

The completion of an accurate and precise map-based genome sequence of *Oryza sativa* cultivar Nipponbare could open the door in understanding genome-wide diversity based on the nucleotide sequence. The genus *Oryza* is composed of 23 species categorized into 10 genome types (Khush 1997). Among them, the wild *Oryza* species represent various closely related species, which show a wide range of diversity. These could provide important resources that could elucidate the evolution of cultivated rice and provide a detailed analysis of divergence in the genus *Oryza*. The present *O. sativa* cultivars have been bred to produce more panicles as well as many of the agronomic characteristics inherent to wild rice, such as tolerance to biotic and abiotic stresses, some of which must have been lost in the process of domestication and breeding. Gene transfer has been successful between species with the AA genome. The rice bacterial blight disease-resistant gene *Xa21* from *O. longistaminata* has been transferred to *O. sativa* by ordinary crossing (Khush et al. 1991) and the resulting progenies have been widely used for breeding new rice varieties resistant to *Xanthomonas* infection (Singh et al. 2001). Other desirable genes conferring traits such as blast resistance or insect resistance have been successfully transferred from a wild species to cultivated rice (Amante-Bordeos et al. 2004). Gene transfer to *O. sativa* from species other than the AA genome has also been possible by embryo rescue technique (Multani et al. 2004).

Rice Biology in the Genomics Era

Hirano, H.-Y.; Hirai, A.; Sano, Y.; Sasaki, T. (Eds.)

2008, XXI, 381 p., Hardcover

ISBN: 978-3-540-74248-7